

APPROXIMATIONS TO DENSITIES IN GEOMETRIC PROBABILITY

BY

HERBERT SOLOMON and MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 24

OCTOBER 27, 1978

PREPARED UNDER GRANT

DAAG29-77-G-0031

FOR THE U.S. ARMY RESEARCH OFFICE

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



Approximations to Densities in Geometric Probability

By

Herbert Solomon and Michael A. Stephens

TECHNICAL REPORT NO. 24

October 27, 1978

Prepared under Grant DAAG29-77-G-0031

For the U.S. Army Research Office

Herbert Solomon, Project Director

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 264.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

1. Introduction.

Many problems in geometric probability involve random variables whose distributions have so far been impossible to find, or, if known, are intractable. Examples of these are the area A , the number of sides N , and the perimeter L of polygons formed by a homogeneous Poisson field of random lines in the plane; or the same variables for Voronoi polygons, which arise when crystals are grown uniformly about points in a plane which have been created by a Poisson process (Crain, 1972). Apart from their intrinsic interest, these polygons often arise in applied models; Crain gives a number of references to applications, and discusses the importance of knowing the densities of the random variables given above. For many of these, low order moments can be found even when the densities are not known, and in this note we suggest a simple approximation to the density using the first three moments, which has been found to work well in practice; where some checks are possible, the results are extremely good. The approximation uses the facts that the typical random variable, say X , is known to be positive, and its density has a steep tail at the lower values, and a long upper tail for higher values of X . The density is therefore like a chi-square density, and statisticians have often approximated X by a random variable Y such that Y/c has χ_p^2 distribution; constants c and p are then found by matching the first two moments of X and Y , though the ability to fit only two parameters often leads to a very crude

approximation. If a third moment can be found, the approximation below can be expected to give much improved results. The method is to approximate X by

$$Y = (cw)^k, \quad (1)$$

where w has the χ_p^2 distribution.

The values of c , p and k are found by equating the corresponding moments of Y to those of X . If the first three moments of X about the origin are μ , μ'_2 , μ'_3 , we have

$$\begin{aligned} \mu &= (2c)^k \Gamma(k+v)/C \\ \mu'_2 &= (2c)^{2k} \Gamma(2k+v)/C \\ \mu'_3 &= (2c)^{3k} \Gamma(3k+v)/C, \end{aligned} \quad (2)$$

where $v = p/2$ and $C = \Gamma(v)$. It is convenient to define

$$\begin{aligned} R_2 &= \mu'_2/\mu^2 = c\Gamma(2k+v)/\{\Gamma(k+v)\}^2 \\ R_3 &= \mu'_3/\mu^3 = c^2\Gamma(3k+v)/\{\Gamma(k+v)\}^3. \end{aligned} \quad (3)$$

In fitting the approximation, R_2 and R_3 are calculated from the moments of X and are used in (3) to solve for k and v ; then c is obtained from the expression for μ . Computer routines are available to perform these operations and then to calculate probabilities of significance points of χ^2 even with noninteger degrees of freedom. For work by hand, significance points for χ^2 with degrees of freedom differing by 0.2 are given in Pearson and Hartley

(1972). Thus significance points and probabilities for X may be very easily approximated, since we have

$$P(X < x) = P(\chi_p^2 < (x^{1/k})/c) . \quad (4)$$

We shall illustrate the approximation on an example where three moments are exactly known, and then provide results for several variables whose higher moments have been found from Monte Carlo studies, but for which there are some theoretical results to provide a test of accuracy for the approximation.

2. Areas of random polygons.

Consider a homogeneous Poisson field of random lines in the plane with intensity parameter τ -- that is, if N_p is the number of random lines whose signed distance, p , to the origin is between $-\frac{x}{2}$ and $\frac{x}{2}$, we can write

$$P[N_p = m] = e^{-\tau x} \frac{(\tau x)^m}{m!} .$$

Of interest is the distribution of the areas of the polygons formed by these random lines in the plane. If A is the area ($0 < A < \infty$), it is known that (Solomon, 1978, Chapter 3)

$$E(A) = \frac{\pi}{\tau^2} , \quad E(A^2) = \frac{1}{2} \frac{\pi^4}{\tau^4} , \quad E(A^3) = \frac{4}{7} \frac{\pi^7}{\tau^6} . \quad (5)$$

The first three moments of X are $\mu = \pi$, $\mu'_2 = \pi^4/2$, and $\mu'_3 = 4\pi^7/7$.

Solving (2), we have $c = 0.70745$, $p = 1.97085$, $k = 1.82035$. If X is approximated by the two-parameter $c\chi_p^2$, we have $c = 6.181$, $p = 0.508$. In Table 1 are given the values of the probability $P(A < x)$, given by these approximations.

It is possible also to fit a Pearson curve to the distribution when three moments and the lower end-point are known. The technique is described and illustrated in Solomon and Stephens (1978). Values given by the Pearson curve approximation are recorded also in Table 1. Finally we include results of a Monte Carlo study made some years ago by Stuart Dufour at Stanford University; 947 polygons were generated by 65 lines, with $\tau = 1$. It can be seen that the Pearson curve and chi-square approximations agree very well in the upper tails, and agree to the accuracy available with the Monte Carlo study. In the lower tail, there is some difference between the approximations, and the three-parameter chi-square approximation is much the closest to the results of the Monte Carlo study. Solomon and Stephens (1978) have found other examples in which this chi-square approximation is very good in the lower tail. Although the upper tail is the one which would most likely be used in statistical testing, a good approximation to the density all along the curve will be required if the density of X is to be used in further calculations, perhaps in combination with other random variables. The Pearson curve will usually have a complicated distributional form, so it will certainly not be as useful in these applications as the three-parameter chi-square approximation.

It is interesting to speculate what $E(A^4)$ might be, by trying to find a pattern to the sequence whose first three terms are given in (5) above. A good candidate would appear to be $E(A^4) = 4\pi^{10}/(5\tau^8)$; for $\tau = 1$ this value is 74918.25. This then gives the fourth central moment $\mu_4 = 55822$, and the kurtosis of the distribution, measured by $\beta_2 = \mu_4/\mu_2^2$, is 37.0. If the $(\chi_p^2)^k$ fit were taken to be accurate, we would have $E(A^4) = 107590$, $\mu_4 = 88494$ and $\beta_2 = 58.7$; the Pearson curve fit gives $\mu_4 = 106833$ and $\beta_2 = 70.8$. Unfortunately, neither approximation strongly supports the guess given above, and the reader is invited to further speculation.

3. Number of sides of random polygons.

The distribution of N , the number of sides of a polygon formed by the above process, has been obtained from an extensive simulation study by Roger Miles (personal communication), and was confirmed in the study made by Dufour. This distribution is of course discrete, with the lowest value $N = 3$, and it is known (see, e.g. Solomon 1978, p. 55) that $E(N) = 4$ and $E(N^2) = \pi^2/2 + 12 = 16.935$. We shall approximate the random variable $X = N - 2.5$ by a continuous distribution, beginning at $X = 0$. The moments of X have been calculated from the Miles' results and for the three-parameter chi-square approximation, the constants

are $c = 1.538$, $p = 1.68$, $k = 0.57$. The discrete probability for $N = 3$ may then be found from the area under the continuous curve between $X = 0$ and $X = 1$, for $N = 4$ by the area between $X = 1$ and $X = 2$, etc.. The results of the Miles simulation, of the chi-square approximation, and of a Pearson curve fit, are shown in Table 2. The results of the approximations are excellent, and both approximations compare very well for the one value which may be obtained analytically i.e. $P(N = 3) = 2 - \pi^2/6 = 0.3551$. The approximations are not claimed to be accurate to the decimal places given in Table 2; these are given simply to make the comparison. The mean of N , from the simulation, is 4.000003, and $E(N^2) = 16.9348$; these agree excellently with the theoretical values given above. It might be pointed out that if one were to rely simply on Monte Carlo studies to obtain accurate estimates of the probabilities, many thousands of polygons would be necessary (Solomon, 1978, p. 55).

4. Perimeter of random polygons.

The density of N above has another application; it can be used to approximate the density of L , the perimeter of a random polygon. Let z be $2L/\pi$. It is known that the density of $z_n = 2L_n/\pi$, where L_n is the perimeter of a random polygon of n sides, has the χ_r^2 distribution, where r is $2(n-2)$. Thus let p_n be the probability $P(N=n)$, and let $f_n(z)$ be the χ_r^2

density with $r = 2(n-2)$; if $f(z)$ is the density of z , we have

$$f(z) = \sum_{n=3}^{\infty} p_n f_n(z) . \quad (6)$$

It follows that

$$\mu'_k = \sum_{n=3}^{\infty} p_n \mu'_{nk} \quad (7)$$

where μ'_k is the k -th moment about the origin of z , and μ'_{nk} is the k -th moment about the origin of the χ^2_r distribution with $r = 2(n-2)$. The values of μ'_{n1} , μ'_{n2} , and μ'_{n3} are respectively r , $2r + r^2$, and $8r + 6r^2 + r^3$; then (7) may be used with the results p_n from Miles' simulations on n , to give the moments of z , and hence those of $L = \pi z/2$. An immediate result of (7) is that $E(z) = E(r) = 2(E(n)-2) = 4$, and it may also be shown that $E(z^2) = 2\pi^2 + 8 = 27.739$. The calculations described above gave $E(z) = 4.00000954$, and $E(z^2) = 27.739$, remarkably accurate results. This accuracy, and the accuracy for N above, is a tribute to the accuracy of the Miles' simulations, and suggests that higher moments will be very accurate also. The next higher moment is $E(z^3) = 265.86$, and these first three moments were used to approximate z by both the χ^2 approximation and the Pearson Curve fit. When the results for z are transformed to results for L we have the values listed in Table 3. The constants in the χ^2 approximation for z are $c = 3.648$, $p = 1.743$, and $k = 0.794$; also since it might be expected to be very accurate we list the fourth moment $E(z^4) = 847.061$, derived from this approximation. For the same

approximation fitted to a constant multiplier of z , like L , the constants p and k do not change, but the constant c_1 for L is related to c for z by

$$\frac{E(L)}{E(z)} = \left(\frac{c_1}{c} \right)^k. \quad (8)$$

Here the left-hand side of (8) is $\pi/2$, and $c_1 = 6.461$. Thus the approximation for L has $c = 6.461$, $p = 1.743$, $k = 0.794$.

Simulation studies directly giving the distribution of L were also made by Dufour; these are the Monte Carlo (M.C.) results in Table 3. It can be seen that the two approximations give very good agreement with the Monte Carlo values. The Monte Carlo results (i.e. $\text{Prob}(L < x)$ for the eight values of x given) were used also to provide estimates of the moments of L ; the first three sample moments about the origin were 6.5625, 71.677, 1031.58, and these give moments for z : 4.178, 29.05, 266.16, to compare with those found above using equation (7). The mean is less accurate than before (recall that $E(z) = 4$), reflecting, no doubt, the difference between the size of the Miles and Dufour simulation studies. However, for interest, the two approximations were fitted also using the direct estimates; results are given, under (2), in Table 3. For the χ^2 fit, $c = 28.73$, $p = 1.14$, and $k = 0.59$. These approximations agree slightly better with the Monte Carlo results, as might be expected, since the moments were calculated directly from them; however, the more extensive simulations which

were used to give the first set of approximations (called (1) in Table 3), and the excellent match of the mean and variance with the theory, suggests that approximations (1) will probably be the better ones.

5. Voronoi polygons.

Similar questions arise in connection with the distributions of statistics associated with Voronoi polygons. For the quantities A , the area of the polygon, L the perimeter, and N the number of sides, only means are known theoretically (Crain 1972). Crain gives results $E(A) = 1/\rho$, $E(L/r) = 1/\sqrt{\rho}$, and $E(N) = 6$, where ρ is the intensity of the Poisson point process generating the "centre-points" of the polygons. We shall assume $\rho = 1$. Crain gives Monte Carlo results for the statistics, using 11000 values of N , and 5000 of s and A , and comments that approximations to the densities will be of considerable use in hypothesis testing in various disciplines. We therefore give the distributions for these statistics using the estimated second and third moments for the approximations. For statistics A and N (Tables 4 and 5) both approximations were used, since the distributions have a chi-square shape, but for L (Table 6), which has a distribution like a normal distribution, only the Pearson curve fit was used. A good fit is obtained with the Monte Carlo results, but again we

emphasise that the moments come from these results also. In fact, Crain's second moment for A (1.24) was used in the fit which is presented here: he refers to an earlier estimate (1.28), and when this was used instead of 1.24, a much worse fit resulted.

The parameters in the chi-square approximations are, for A :

$c = 0.723$, $p = 1.855$, $k = 0.445$, and for N : $c = 4.457$, $p = 3.429$, $k = 0.485$.

6. Comments.

The 3-parameter chi-square approximation, and the Pearson curve approximation using a known lower endpoint and three moments, have been fitted to statistics of essentially two types; for the first, such as A and N for random polygons, either theoretical results for the moments were known exactly, or such a large number of Monte Carlo studies had been made that the density could be regarded as giving exact moments; while for the second type of statistic, especially those for the Voronoi polygons, the results for moments other than the mean were found from relatively small Monte Carlo studies. In the first group, and especially when the moments are exactly known theoretically, we can expect the approximations to give excellent results to the densities. For the second group, we have demonstrated that one gets an excellent approximation to the existing Monte Carlo results, indicating that if far more of

these results were available, a simple approximation (the three-parameter chi-square approximation) exists for the random variable. The Pearson curve fit has been included because the agreement between the two approximations tends to give one confidence that both are very good, especially in the long upper tail. However our main purpose has been to suggest the use of the three-moment chi-square approximation, because of its much greater flexibility; it will be especially more useful, if the density of the statistic is to be introduced into further calculations.

Acknowledgements. This work has been supported by the National Research Council of Canada, and by the U. S. Office of Naval Research. The authors wish to thank Mr. C.S. Davis for his considerable help with the computations.

Reference

- Crain, Ian K. (1972) Monte Carlo Simulation of the Random Voronoi Polygons: Preliminary Results. Search, Vol. 3, 220: Australian & New Zealand A.A.S.
- Solomon, Herbert (1978) Geometric Probability, Monograph 28, Regional Conference Series in Applied Mathematics, S.I.A.M.: Philadelphia.
- Solomon, Herbert and Stephens, Michael A. (1978) Approximations to Density Functions Using Pearson Curves, Jour. Amer. Statist. Assn., 73, 150-160.

TABLE 1

Approximations to the distribution of A , the area of a random polygon.

The table entries are $P(A < x)$, for the various approximations.

P.C. = Pearson curve with lower end-point fixed; M.C. Monte Carlo study.

Approximation:	x:	0.05	0.10	0.25	0.50	0.75	1.0	2.5	5.0	10.0	15.0
$c\chi_p^2$:	.272	.324	.408	.485	.535	.574	.707	.813	.907	
P_c	:	.203	.255	.345	.431	.489	.534	.693	.815	.916	.956
$(c\chi_p^2)^k$:	.132	.186	.288	.390	.460	.514	.695	.823	.920	.957
M.C.	:	.13	.18	.27	.38	.45	.50	.67	.80	.90	.95

TABLE 2

Approximations to the distribution of N, the number of sides of random polygons

The table gives values $P(N = n)$, by simulation (M.C.) and two approximations.

Approximation	n =	3	4	5	6	7	8	9	10
M.C. *	:	.355	.381	.190	.059	.013	.002	.0003	.00003
$(c\chi_p^2)^k$:	.358	.374	.189	.063	.015	.002	.0003	.00003
P.C.	:	.353	.377	.191	.061	.016	.002	.0002	.00002

* The exact value for $n = 3$ is 0.355066.

TABLE 3

Approximations to the distribution of L, the perimeter of a random polygon.

The table entries are $P(L < x)$. M.C. refers to Monte Carlo results (see Section 4). Approximation (1) uses moments calculated from equation (7) and M.C. results for N; approximation (2) uses moments for L calculated directly from M.C. results.

Method	x:	0.5	1.0	2.5	5.0	7.5	10.0	15.0	20.0
M.C.		.05	.11	.26	.51	.67	.79	.92	.98
(1) {	$(c\chi_p^2)^k$.057	.110	.261	.478	.649	.775	.919	.976
	P.C.	.046	.100	.257	.479	.643	.775	.920	.975
(2) {	$(c\chi_p^2)^k$.052	.109	.277	.511	.682	.799	.925	.974
	P.C.	.046	.103	.271	.512	.674			

TABLE 4

Approximations to the distribution of A, the area of Voronoi polygons

The table gives values of $P(A < x)$.

Approximation	x:	.1	.2	.3	.4	1.0	1.2	1.5	2.0
$(c\chi_p^2)^k$.006	.025	.058	.161	.535	.678	.840	.968
P.C.		.004	.020	.052	.155	.538	.675	.838	.967

TABLE 5

Approximations to the distribution of N, the number of sides of
Voronoi polygons.

The table entries are $P(N = n)$.

n:	3	4	5	6	7	8	9	10	11	12
$(cx_p^2)^k$.014	.117	.253	.283	.198	.094	.032	.0076	.0013	.0002
P.C.	.012	.116	.256	.280	.200	.096	.036	.0071	.0010	.0001
M.C.	.011	.110	.259	.288	.206	.087	.029	.0077	.0014	.0002

TABLE 6

Approximations to the distribution of $u = L/4$, where L is the perimeter of
Voronoi polygons.

The table entries are $P(u < x)$.

x:	.1	.2	.3	.4	.5	.6	1.0	1.2	1.5
P.C.	.0024	.0048	.0093	.0187	.0346	.0604	.4586	.7120	.9959

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 24	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Approximations to Densities in Geometric Probability		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Herbert Solomon and Michael A. Stephens		8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-G-0031
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-14435-M
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE October 27, 1978
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release; Distribution Unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents. This report partially supported under Office of Naval Research Contract N00014-76-C-0475 (NR-042-267) and issued as Technical Report No. 264.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Approximating densities, chi-square approximations, geometric probability, random polygons, Voronoi polygons.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Please see reverse side.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 63 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Many random variables arising in problems of geometric probability have intractable densities, and it is very difficult to find probabilities or percentage points based on these densities. A simple approximation, a generalization of the chi-square distribution, is suggested. To approximate such densities; the approximation uses the first three moments. These may be theoretically derived, or may be obtained from Monte Carlo sampling.

The approximation is illustrated on random variables (the area, the perimeter, and the number of sides) associated with random polygons arising from two processes in the plane. Where it can be checked theoretically, the approximation gives good results. It is compared also with Pearson curve fits to the densities.